

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Express Mail Label No.: EV287336791US

Attorney Docket No. 21465-501 CIP2

Date of Deposit: April 23, 2004



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS:	Jonathan M. Rothberg, et al.		
ASSIGNEE:	CuraGen Corporation		
SERIAL NUMBER:	09/814,338	EXAMINER:	Young J. Kim
FILING DATE:	March 21, 2001	ART UNIT:	1637
FOR:	METHOD OF SEQUENCING A NUCLEIC ACID		

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF MARCEL MARGULIES, Ph.D. UNDER 37 C.F.R. §1.132

I, MARCEL MARGULIES, declare and state that:

1. I am Vice President of Engineering, at 454® Life Sciences, the exclusive licensee of this application. My previous employment includes Director of New Technology Research at Perkin-Elmer's Instrument Division in Norwalk, CT, and Associate Director of the Hubble Space Telescope project.
2. I earned my B.Sc. in Engineering from the Free University of Brussels, in Belgium, and a Ph.D. in theoretical physics from Columbia University.
3. I have reviewed the instant application and the November 6, 2003 Office Action in this case.
4. Based on information and belief, it is my opinion that the claimed sequencing apparatus and substrate of the instant application are vastly superior to other sequencing systems, including the system reported by Chee et al. in Published Application No. U.S. 2003/0108867 ("Chee et al."), as cited by the Examiner.

The superior performance of the claimed invention can be attributed to the functional features of the apparatus and substrate, which include: 1) compact wafers; 2) attachable optical fibers; 3) flow chamber and fluid means; and 4) specific fiber and well sizes. As a result of these features (and others), the claimed invention provides the first massively parallel, solid-phase, whole-genome sequencing platform that can be scaled for viral, bacterial, and even human genomes.

Functional features of the claimed substrate and apparatus:

compact wafers can be placed in flow chambers for efficient fluid exchange

5. As recited in the claims, the compact wafers of the invention allow *placement into flow chambers*, which utilize an *efficient fluid exchange system* and thereby provide significantly faster sequence analysis. The advantages of the claimed apparatus and substrate are fully disclosed in the instant application, as filed.
6. The instant application teaches that the claimed apparatus and substrate include a compact wafer formed from a bundle of optical fibers, cut and polished to a thickness of 0.5 mm to 5 mm. This teaching is found, *inter alia*, in the originally filed application on p. 36, l. 12-15; and p. 36, l. 30 to p. 37, l. 3.
7. The instant application additionally teaches that the claimed substrate can be used with a flow chamber and fluid means for delivering sequencing reagents and washes to the wafer surface. This teaching is found, *inter alia*, in the originally filed application in Figs. 2 and 3; on p. 4, l. 14-18; p. 30, l. 15-19; p. 33, l. 20 to p. 35, l. 20; and in Exmp. 3 on p. 53, l. 28 to p. 54, l. 15.
8. The instant application further teaches that the underside of the compact wafer in the flow chamber can be optically linked or directly contacted with a fiber optic bundle to allow image capture, for example, through a CCD system. This teaching is found, *inter alia*, in the originally filed application on p. 34, l. 13-18; and Fig. 2.

9. The instant application teaches also that the underside of the compact wafer in the flow chamber can be placed in proximity to conventional optics mechanism, e.g., a high numerical aperture lens system to allow for image capture. This teaching is found, *inter alia*, in the originally filed application on p. 34, l. 19-23.
10. The claimed wafer for use in the flow chamber allows for much faster sequence analysis. The claimed apparatus and substrate thereby yield significantly improved results, which are not obtained with other sequencing systems such as that reported by Chee et al.
11. As cited by the Examiner, Chee et al. does not specify the length of the optic fibers for their sequencing system. Instead, Chee et al. relies on WO 98/50782 (see Chee et al., ¶ [0007]), which reports the use of optic fibers that are several meters long (Ex. 1). These long and bulky fibers cannot readily fit into a flow chamber. Instead, Chee et al. reports methods of inverting the long optic fiber and sequentially dipping the tip into individual cups filled with solutions of single nucleotides (Chee et al., ¶¶ [0192] – [0195], *inter alia*). This awkward dipping process is necessitated by the long, bulky fiber employed in Chee et al.
12. The dipping method of Chee et al. is predicted to be completely or partly inoperable. First, the continuous plunging of an optic fiber tip into nucleotide solutions would tend to dislodge any beads from the wells. Second, for DNA attached directly to wells, dipping would be ineffective in delivering the nucleotide solution to the wells due to the counteraction of air pressure. This phenomenon is generally observed for inverted cups or glasses placed into reservoirs of water, and provides the basis for the oceanographic apparatus known as the diving bell. An illustration provided as an aid in understanding is included as Ex. 2 (adapted from <http://home.earthlink.net/~dmocarski/chapters/chapter7/main.htm>).

13. Even assuming, *arguendo*, that the dipping method of Chee et al. were marginally effective in delivering reagents and preserving samples for sequencing, the method would be extremely slow and inefficient compared to the claimed invention. The long fiber optic tip of Chee et al. would need to be inverted and dipped in and out of at least four cups (e.g., A, T, C, G), or perhaps more than eight cups (e.g., A, first wash, T, second wash, C, third wash, G, fourth wash) to determine only one nucleotide of sequence. In contrast, compact wafer of the invention can be placed into a flow chamber to allow for rapid and efficient delivery of sequencing reagents and washes to the compact wafer. An illustration provided as an aid in understanding is included as Ex. 3.
14. Because the long optic fibers of Chee et al. cannot readily fit into a flow chamber, the sequencing reactions are performed using an unwieldy dipping process. This results in significant delays and increased sample losses. By comparison, the claimed wafer is fitted into a flow chamber to allow streamlined processing and sequence analysis. This is a *significant functional advantage* over the system reported in Chee et al.

Functional features of the claimed substrate and apparatus:

compact wafers have well sizes to maximize signal capture and minimize sample loss

15. As recited in the claims, the compact wafers of the invention include *optimally sized fibers and wells* that allow *maximal signal capture and minimal sample loss* and thereby provide significantly improved sequence analysis. The advantages of the claimed apparatus and substrate are fully disclosed in the instant application, as filed.
16. The instant application teaches that the claimed apparatus and substrate include a compact wafer that includes optic fibers with a diameter of 3 μm to 100 μm . The application teaches that this diameter is important to ensure that each light signal can be captured as a single pixel. This teaching is found, *inter alia*, in the originally filed application on p. 36, l. 15 and 25-29.

17. The instant application teaches that the claimed apparatus and substrate include a compact wafer that includes wells with a depth of one-half to three times the diameter of the optic fibers. The application expressly recognizes the problem of bead/sample loss during the sequencing reaction. This teaching is found, *inter alia*, in the originally filed application on p. 37, l. 6-9; p. 39, l. 22-24; and Fig. 4.
18. The claimed fiber diameter and well depth of the compact wafer allow for much more effective sequence analysis. Accordingly, the claimed apparatus and substrate yield significantly improved results, which are not obtained with other sequencing systems, including the system reported by Chee et al.
19. As cited by the Examiner, Chee et al. ¶ [0105] apparently reports the use of long optic fibers with diameters ranging from approximately 0.17 μ m to 0.03 μ m. These diameters can be calculated from “high density” arrays indicated by Chee et al. ¶ [0105], i.e., arrays containing 40,000 fibers/mm² to 1,000,000 fibers/mm² (Office Action, quoting Chee et al. on pages 9-10). Chee et al. appears to be silent as to the specific well depths employed with the optic fibers.
20. With diameters of approximately 0.17 μ m to 0.03 μ m, many of the optic fibers employed by Chee et al. would produce sequencing systems that are completely or partly inoperable. Optic fibers having a such small diameters would require bead and well sizes less than 0.17 μ m to 0.03 μ m in diameter. Such systems would be predicted to have a myriad of problems, including difficulties in distinguishing light signals from each fiber and in depositing the beads in the wells. By comparison to Chee et al. ¶ [0105], the compact wafer of the invention employs optic fibers 3 μ m to 100 μ m in diameter to provide for maximal sample density while still allowing accurate signal detection and efficient bead delivery.
21. Chee et al. do not appear to specify well depths for use with the optic fibers, and evidently fail to recognize the importance of well depth in preventing sample loss.

In fact, Fig. 1 in Chee et al. shows beads and samples jutting out from their wells. This configuration would likely lead to significant sample loss during the “invert and dip” process reported by Chee et al. In contrast, the compact wafer of the invention employs well depths of one-half to three times the diameter of the fiber, which are important in minimizing sample loss during preparation and analysis. The optimally sized fibers and wells (¶¶ 16 and 17, above) therefore represent *significant functional advantages* over the system reported in Chee et al.

**Superior function of the claimed substrate and apparatus:
massively parallel analysis of viral and human genomic sequences**

22. As a result of these highly advantageous, functional features (¶¶ 14, 20, and 21, above), and other important aspects, the substrate and apparatus claimed in the instant application are the first to allow rapid massively parallel sequencing for whole genomes.
23. Traditional methods for genome sequencing have been slow, expensive, laborious, and industrial-scale, since they involve individually preparing and sequencing DNA fragments of the genome. The Human Genome Project, for example, required approximately 12 years, \$2.7 billion dollars, and 60 million samples to complete.
24. In contrast, the substrate and apparatus claimed in the instant application provide a massively parallel, scalable platform that dramatically reduces the time, cost, sample preparation, and space required for genome sequencing. Instead of individually preparing and sequencing each sample, the claimed substrate and apparatus allow parallel sequencing of thousands (or hundreds of thousands) of samples.
25. Recently, the claimed substrate and apparatus were used to sequence the entire adenovirus genome (approximately 30,000 base pairs) contained on an expression vector in less than one day (see NY Times article, Ex. 4). The entire sequencing

process from sample preparation to data analysis was accomplished in less than one day, and provided over 99% genome coverage. The resulting adenovirus sequence was published in GenBank under Accession Nos. AY370909, AY370910, and AY370911 (Ex. 5).

26. In further experiments, the apparatus of the instant application was used to sequence a segment human chromosome 12 (approximately 170,000 base pairs) contained on an artificial chromosome vector (Ex. 6). With the apparatus, a one-day sequencing run produced sufficient shotgun sequence coverage of the chromosome 12 clone (Ex. 6, p. 6). A single sequencing run obtained 85% genome coverage and 98% consensus accuracy (Ex. 6, p. 3). These results were presented at the 15th Annual Genome Sequencing and Analysis Conference, held on September 21-24, 2003 (Ex. 6, p. 1).
27. To generate this sequence information described in ¶¶ 25 and 26 (above), preferred commercial embodiments of the claimed substrate and apparatus were fabricated. In these preferred embodiments, the claimed substrates (termed "PicoTiter Plates") were formed from cavitated fiber optic wafers formed from a fused bundle of a plurality of individual optical fibers as taught and claimed by the instant application.
28. Specifically, PicoTiter Plates were made acid etching the top surface of fiber optic wafers to form wells with diameters between 39 and 44 μm , as currently claimed. The fiber optic wafer exhibited a thickness of about 2.0 mm, also as currently claimed. In addition, the wells on PicoTiter Plates were fabricated with depths ranging from 26 to 76 μm (i.e., from between one half the diameter of an individual optical fiber and three times the diameter of an individual optical fiber, as recited in the pending claims). Finally, the wells were loaded with nucleic acid template and beads with pyrophosphate sequencing reagents attached thereto, as recited in the pending claims. Sequencing by synthesis was then performed as

described in the specification, and using the claimed apparatus to flow sequencing reagents over the PicoTiter Plate.

29. The substrate and apparatus claimed in the instant application therefore fulfill a long-felt but unmet need for rapid, whole-genome analysis of viral and bacterial pathogens (e.g., ¶ 25, above). Such analysis is critical for biodefense, drug discovery, and the identification of emerging pathogens. More than this, the claimed apparatus solves the long-standing problems with analysis of large genomes, such as human genomes (e.g., ¶ 26, above). Solutions for large-genome sequencing are vital for drug development, early diagnosis, and faster clinical interventions.
30. For these reasons, in my opinion, the claimed substrate and apparatus represent a significant advancement in the field as the first massively parallel, solid-phase, whole-genome sequencing platform that can be scaled for the smallest to the largest genomes.

Conclusion

31. Therefore, based on information and belief, and all of the foregoing, it is my opinion the claimed sequencing apparatus and substrate substantially outperform the sequencing platforms used by Chee et al. and others. This is due to the functionally superior features of the claimed invention, which include compact fiber optic wafers, detachable fiber optic bundles, flow chambers and fluid means, and specifically sized fibers and wells. All of these features, and the other aspects of the invention, work together to achieve significantly faster results compared to other sequencing systems.
32. I declare that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18

Applicants: Rothberg, et al.
U.S.S.N. 09/814,338

U.S.C. § 1001 and that willful false statements may jeopardize the validity of this application and any patent issuing therefrom.

Dated: 4/20/04

Signed: 
MARCEL MARGULIES

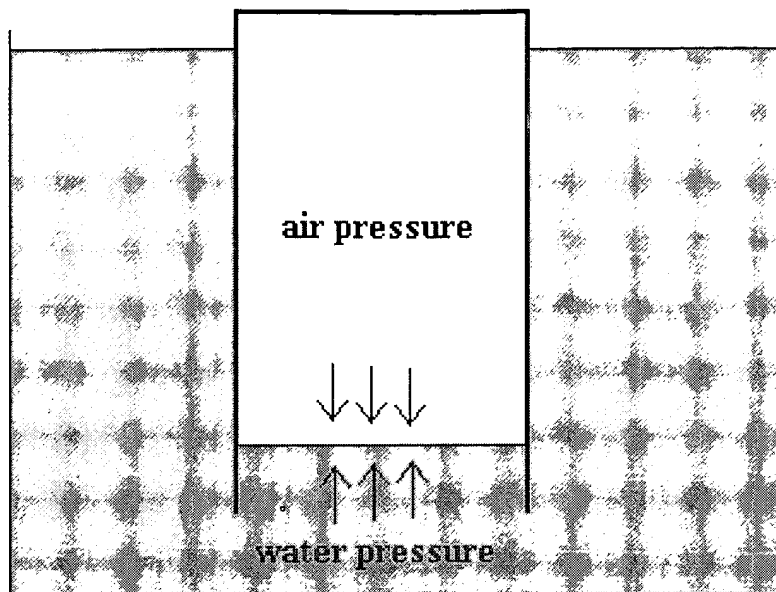
A. The Format Choices

The individually clad, optical fiber strand

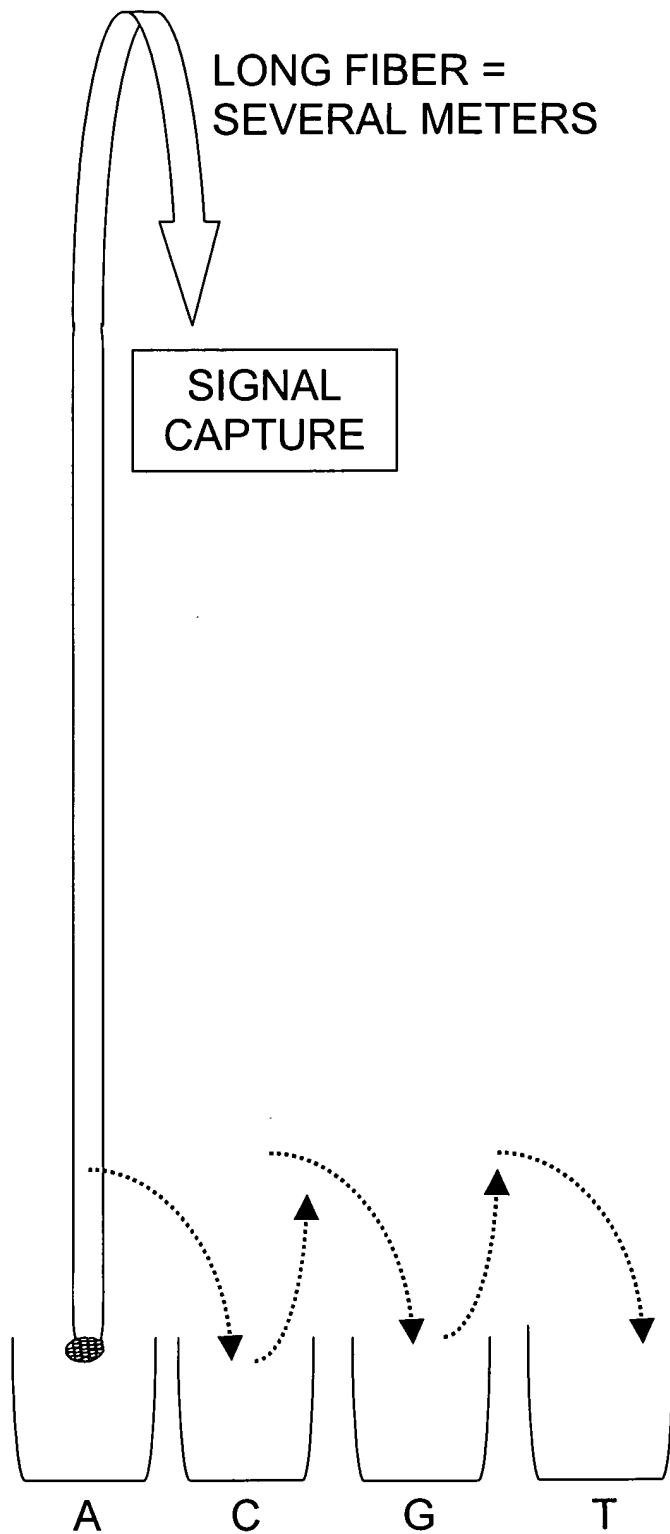
A typical optical fiber strand is illustrated by Figs. 1 and 2A and 2B. As seen therein, an individual optical fiber strand 10 is comprised of a single optical fiber 12 having a rod-like shaft 14 and two fiber ends 16, 18, each of which provides a substantially planar end surface. The intended distal surface 20 at the fiber end 16 is illustrated by Fig. 2A, while the intended proximal surface 22 at the fiber end 18 is illustrated within Fig. 2B. It will be recognized and appreciated that the terms "proximal" and "distal" are relative and interchangeable until the strand is ultimately positioned in an apparatus. The optical fiber 12 is composed typically of glass or plastic; and is a flexible rod able to convey light energy introduced at either of its ends 16 and 18. Such optical fibers 12 are conventionally known and commercially available. Alternatively, the user may himself prepare individual optical fibers in accordance with the practices and techniques reported in the scientific and industrial literature. Accordingly, the optical fiber 12 is deemed to be conventionally known and available as such.

It will be appreciated that Figs. 1-2 are illustrations in which the features have been purposely magnified and exaggerated beyond their normal scale in order to provide both clarity and extreme detail. Typically, the conventional optical fiber has a cross section diameter of 5-500 micrometers; and is routinely employed in lengths ranging between meters (in the laboratory) to kilometers (in field telecommunications). Moreover, although the optical fiber 12 is illustrated via Figs. 1-2 as a cylindrical extended rod having substantially circular proximal and distal end surfaces, there is no requirement or demand that this specific configuration be maintained. To the contrary, the optical fiber may be polygonal or asymmetrically shaped along its length; provided with special patterns and shapes at the proximal and/or distal faces; and need not present an end surface which is substantially planar. Nevertheless, for best efforts, it is presently believed that the substantially cylindrical rod-like optical fiber having planar end surfaces is most desirable.

Each optical fiber 12 is desirably, but not necessarily, individually clad axially along its length by cladding 26. This cladding 26 is composed of any material with a lower refractive index than the fiber core and prevents the transmission of light energy photons from the optical fiber 12 to the external environment. The cladding material 26 may thus be composed of a variety of radically different chemical formulations including various glasses, silicones, plastics, platings, and shielding matter of diverse chemical composition



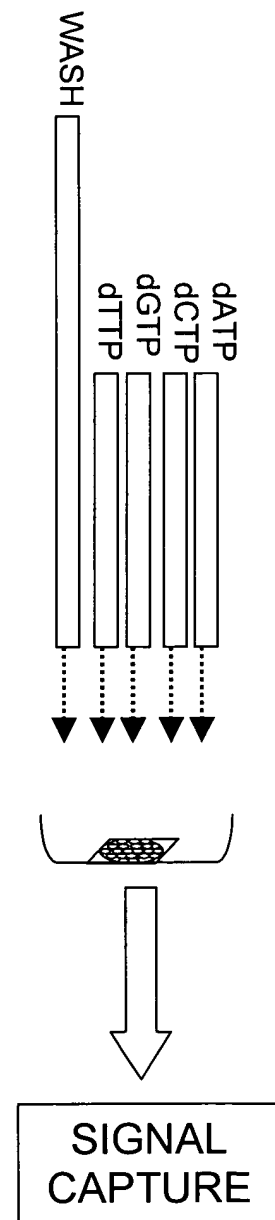
CHEE et al.



CANNOT USE FLOW CHAMBER –
REQUIRES SLOW, WASTEFUL
DIPPING PROCESS

CLAIMED SUBSTRATE

COMPACT WAFER =
0.5 mm to 5 mm thick



CAN USE FLOW CHAMBER –
ALLOWS FAST & EFFICIENT
PROCESSING

Home - Site Index - Site Search/Archive - Help

Go to a Section



Quotes:



Welcome, ebaetge - Member Center - Log Out

Search:



The New York Times
global economies
A SPECIAL ADVERTISING SECTION

A Hot Tip from the World of
International Business

NYTimes.com > Business

Company Says It Mapped Genes of Virus in One Day

By ANDREW POLLACK

A small company developing a novel method of sequencing genes said yesterday that it had determined the entire genetic code of a virus in a single day.

The company, 454 Life Sciences, said it was the first time that the entire genome of an organism had been sequenced using an unconventional technique and demonstrated the feasibility of its technology, which it said might eventually allow gene sequencing to be done faster and less expensively.

"It's a real threshold moment," said Richard Gibbs, director of the genome-sequencing center at Baylor College of Medicine in Houston and an adviser to the company. "This is going to be big."

Other scientists were more skeptical, saying that sequencing a virus, which has a tiny genome, is trivial and that there are aspects of the

Advertisement

Real Estate

Sign up for E-Mail Alerts!
Receive the latest property listings in your inbox...

ARTICLE TOOLS

☒ E-Mail This Article

☐ Printer-Friendly Format

☐ Most E-Mailed Articles

☐ Reprints & Permissions

ARTICLE TOOLS
SPONSORED BY

STARBUCKS.COM

TIMES NEWS TRACKER

Topics

Viruses

Alerts

Create

Genetics and Heredity

Create

Baylor College of Medicine

Create

Respiratory Diseases

Create

Create Your Own

Manage

Most Popular Alerts

Take a Tour

CLICK HERE TO SUBSCRIBE

technology that might make it difficult to do more complex organisms.

"I think doing a whole bacterium will be a challenge," said Edward M. Rubin, director of the Joint Genome Institute, a Department of Energy sequencing center in Walnut Creek, Calif. Bacteria are the next level up in complexity after viruses.

454, based in Branford, Conn., is one of several companies racing to improve gene

sequencing. Some scientists say it might one day be possible to sequence an entire human genome in a few days for as little as \$1,000 so that each person could have his or her genome for use in predicting susceptibility to disease and choosing the most appropriate medicines. The Human Genome Project, which first sequenced a human genome, took years and cost tens of millions of dollars.

Still, the thousand-dollar genome is years away. The human genome, which consists of more than three billion letters of the genetic code, is about 100,000 times the size of the adenovirus that 454 sequenced, which has a genome of about 33,000 letters.

454 is majority owned by CuraGen, a genomics company. The name 454 was the code name by which the project was referred to at CuraGen and the numbers have no special meaning, Richard F. Begley, chief executive of 454, said. Some of the company's technology is licensed from Pyrosequencing A.B., a Swedish company. Mr. Begley said the company hoped to start offering a sequencing service by the end of the year and to begin selling machines about six months after that.

Sequencing is usually done in big sequencing centers with dozens of machines and dozens of people. Dr. Begley said 454's technology would pave the way for the "personalized genome center" in which a single scientist could do sequencing on one

FISHER
INVESTMENTS

Has a new bull market
just started?

Ken Fisher says "YES!"

"This is a winning investment climate!"

Get his **NEW Stock Market Outlook**
absolutely FREE!



 **Click Here**

Forbes columnist & money manager, Ken Fisher

Sell or Rent Your Home
Post a property listing on
NYTimes.com...

Find a Mover
Get instant quotes for full-
service, self-service and last-
minute moves...

Get Mortgage Quotes
Get instant mortgage quotes
and calculate payments...

machine in a small room.

The company's technique does more sequencing in parallel than the conventional technique, he said. Another significant time-saving development, he said, comes in preparing the samples for sequencing. The 454 technique, he said, requires one sample preparation for each organism compared with dozens and even thousands for the conventional technique.

Scientists differed on whether sequencing a virus in a day was truly fast. The Department of Energy once did 30 bacteria in 30 days. The virus that causes SARS was sequenced in six days, but it might be misleading to compare that effort with 454's.



Expect the World every morning with home delivery of The New York Times newspaper.

Get The Times for as little as \$2.90 a week.

Advertiser Links

[\\$7 Trades at Scottrade](#)

[Free Portfolio Builder demo from OppenheimerFunds](#)

[Click here to get your FREE 2003 Stock Market Outlook](#)

RELATED ARTICLES

- [A Better Mousetrap Is Built, Detecting Deadly Bacteria](#) (July 11, 2003) \$
 - [Specter Asks Bush to Permit More Embryonic Cell Lines](#) (April 23, 2003) \$
 - [PUBLIC LIVES: A Lab of Dying Mice, and a Virus in His Recipe Box](#) (July 19, 2002) \$
 - [TRACES OF TERROR: THE SCIENCE: SCIENTISTS CREATE A LIVE POLIO VIRUS](#) (July 12, 2002) \$
 - [THE YEAR IN IDEAS: A TO Z: Attaching Good Genes to Bad Viruses](#) (December 9, 2001) \$
- Find more results for [Viruses](#) and [Genetics](#) and [Heredity](#).

TOP BUSINESS ARTICLES

- [Schering-Plough Details Its Steps for a Turnaround](#)
- [Ouster at Freddie Mac Discussed](#)
- [Stocks Surge After Intel Lifts Guidance](#)
- [Utilities Divided Over Managing Power Flow](#)
- [Abbott Spins Off Hospital Products Unit](#)



CGCTCAGGATACGACCTTCGCTCCTAGATCGGATCCCCGGGCTTATTATATAGCTCGATCGATCT
TTCTCTATATCCGCGCTGCTGCTATATACACACACACCGCGGATAGCATGACTGATCTA
CCCCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CACAGACTTACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

PubMed

Nucleotide

Protein

Genome

Structure

PMC

Taxonomy

OMIM

Book

Search for

Limits

Preview/Index

History

Clipboard

Details

Show: ☐ 1: AH013174. Expression vector...[gi:34014916][Links](#)

LOCUS AY370909S1 2617 bp DNA linear SYN 21-AUG-2003

DEFINITION Expression vector pAdEasy-1, contig 1.

ACCESSION AY370909

VERSION AY370909.1 GI:34014917

KEYWORDS .

SEGMENT 1 of 3

SOURCE Expression vector pAdEasy-1

ORGANISM Expression vector pAdEasy-1
artificial sequences; vectors.

REFERENCE 1 (bases 1 to 2617)

AUTHORS Sarkis,G., Costa,G., Leamon,J., Maithreyan,S., Berka,J., Du,L.,
Fierro,J., McDade,K., Puc,B., Roth,G.T., Gomes,X., Altman,W.,
Charumilind,J., Chen,Y.-J., Chen,Z., de Winter,A., Dewell,S.,
Drake,J., Forte,R., He,W., Helgesen,S., Jannotti,M.L., Jarvie,T.,
Jirage,K., Kelch,K., Kim,J.-B., Kukanski,K., Lanza,J., Lee,W.,
Lefkowitz,S., Lu,H., Makhijani,V., Margulies,M., Nobile,J.,
Norton,W., Reifler,M., Rodgers,G., Ronan,M., Simpson,J.,
Tartaro,K., Verma,S., Zimmerman,Z., Dacey,P., Begley,R. and
Lohman,K.TITLE Sequence Analysis of the pAdEasy-1 Recombinant Adenoviral Construct
Using the 454 Life Sciences Sequencing-by-Synthesis Method

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2617)

AUTHORS Lohman,K.

TITLE Direct Submission

JOURNAL Submitted (18-AUG-2003) 454 Life Sciences, 20 Commercial Street,
Branford, CT 06405, USA

FEATURES Location/Qualifiers

source

1..2617

/organism="Expression vector pAdEasy-1"

/mol_type="other DNA"

/db_xref="taxon:243021"

/clone="Stratagene catalog number 240005"

/note="contig 1; differs from pAdEasy-1 sequence from
Stratagene; sequenced by new method"

BASE COUNT 576 a 748 c 711 g 582 t

ORIGIN

```
1 aattaacatg catggatcct acgtctcgac cgatgccctt gagagccttc aacccagtca
61 gctccttccg gtgggcggcg gggcatgact atcgctcgccg cacttatgac tgtcttcttt
121 atcatgcaac tcgtaggaca ggtgccggca gcgctctggg tcattttcgg cgaggaccgc
181 tttcgctgga gcgcgacgat gatcggcctg tcgcttgccg tattcggaat cttgcacgcc
241 ctcgctcaag ctttcgtcac tgggtcccgcc accaaacgtt tcggcgagaa gcaggccatt
301 atcgccggca tggcgggcca cgcgctggc tacgtcttgc tggcggttcgc gacgcgaggc
361 tggatggcct tccccattat gattcttctc gcttccggcg gcacgaggat gccgcgcttg
421 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgctc
481 gcggctctta ccagcctaac ttcgatcatt gttggaccgc tgatcgtcac ggcgatttat
541 gccgcctcgg cgagcacatg gaacgggttg gcatggattg taggcgccgc cctatacctt
```

```

601 gtctgcctcc ccgcgttgcg tcgcggtgca tggagccggg ccacctcgac ctgaatggaa
661 gccggcgcca cctcgctaac ggattcacca ctccaagaat tggagccaat caattcttgc
721 ggagaactgt gaatgcgcaa accaaccctt ggcagaacat atccatcgcg tccgccatct
781 ccagcagccg caccgcgccg atctcgggca gcggttgggtc ctggccaacg gtgcgcacatga
841 tcgtgctcct gtcgttgagg acccggttag gctggcgggg ttgccttact ggtagcaga
901 atgaatcacc gatacgcgag cgaacgtgaa gcgactgctg ctgcaaacg tctgcgacct
961 gagcaacaac atgaatggtc ttcggtttcc gttgtttcgt aaagtctgga aacgcggaag
1021 tcagcgccct gcaccattat gttccggatc tgcacgcag gatgctgctg gctaccctgt
1081 ggaacaccta catctgtatt aacgaagcgc tggcattgac cctgagtgat ttttcttctg
1141 gtcccgcgcg atccataacc ccagttgttt accctcaca cgttcagta accgggcatg
1201 ttcatcatca gtaaccgcgc atcgtgagca tctctctctg tttcatcggt atcattaccc
1261 ccatgaacaa gaaatcccc ttacacggag gcatcagtga ccaaacaagg aaaaaaccag
1321 cccttaacat ggcccgcttt atcagaagcc agacattaac gcttctggag aaactcaacg
1381 agctggacgc ggaatgaacg gcagacatct gtgaatcgct tcacgaccac gctgatgagc
1441 tttaccgcag ctgcctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc
1501 tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgctcagg
1561 gcgcgtcagc ggggtgttggc ggggtgtcggg gcgcagccat gaccagtcga cgtagcgata
1621 gcggcagtgat atactggctt aactatgcgg catcagagca gattgtactg agagtgcacc
1681 atatgcggtg tgaataaccg cacagatgcg taaggaagaa aataaccgca tcaggcgctc
1741 ttccgcttcc tcgctcactg actcgtcgcg ctccggtcgtt cggctgcggc gagcggtatc
1801 agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa
1861 catgtgagca aaaggccaag caaaggcca aggaaccagt aaaaaggccg cgttgctggc
1921 gtttttccat aggtcccgcc cccctgacga gcatcacaaa atcgacgctc aagtcagagg
1981 tggcgaaacc cgacaggact ataaagatac caggcgtttc ccctggaagc tccctcgtgc
2041 gctctcctgt taccgaccct gccgcttacc ggatacctgt ccgcctttct tcccttcggg
2101 aagcgtggcg ctttctcata agctcacgct gtaggtatct cagttcgggt taggtcgttc
2161 gtcceaagct gggctgtgtg caccgaaccc cgttcagcc cgaccgctgc gccttatccg
2221 gtaactatcg tcttgagtc aaccgggtaa gacacgactt atcgccactg gcagcagcca
2281 ctggtaacag gattagcaga gcgaggtatg taggcgggtc tacagagttc ttgaagtggg
2341 ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag
2401 ttaccttcgg aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg
2461 tggttttttg tttgcaagca gcagattacg cgcaagaaaa aaggaatctc aagaagattc
2521 ctttgtattc ttttcttacg ggggtgctgac gctcagtgga acgaaaactc acgttaaggg
2581 attttggtca tgagattatc aaaaaggatc ttcacct

```

//

```

LOCUS      AY370909S2              1062 bp    DNA        linear    SYN 21-AUG-2003
DEFINITION Expression vector pAdEasy-1, contig 2.
ACCESSION  AY370910
VERSION    AY370910.1  GI:34014918
KEYWORDS   .
SEGMENT    2 of 3
SOURCE     Expression vector pAdEasy-1
  ORGANISM Expression vector pAdEasy-1
            artificial sequences; vectors.
REFERENCE  1 (bases 1 to 1062)
  AUTHORS  Sarkis,G., Costa,G., Leamon,J., Maithreyan,S., Berka,J., Du,L.,
            Fierro,J., McDade,K., Puc,B., Roth,G.T., Gomes,X., Altman,W.,
            Charumilind,J., Chen,Y.-J., Chen,Z., de Winter,A., Dewell,S.,
            Drake,J., Forte,R., He,W., Helgesen,S., Jannotti,M.L., Jarvie,T.,
            Jirage,K., Kelch,K., Kim,J.-B., Kukanski,K., Lanza,J., Lee,W.,
            Lefkowitz,S., Lu,H., Makhijani,V., Margulies,M., Nobile,J.,
            Norton,W., Reifler,M., Rodgers,G., Ronan,M., Simpson,J.,
            Tartaro,K., Verma,S., Zimmerman,Z., Dacey,P., Begley,R. and
            Lohman,K.
  TITLE    Sequence Analysis of the pAdEasy-1 Recombinant Adenoviral Construct
            Using the 454 Life Sciences Sequencing-by-Synthesis Method
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 1062)
  AUTHORS  Lohman,K.
  TITLE    Direct Submission

```

```
3121 atccttgcaa gtctagcgcc tgctgccatg cgcgggcggc aagcgcgcgc tcgtatgggt
3181 tgaggtgggg gaccccatgg catgggggtg gtgagcgcgg aggcgtacat gccgcaaattg
3241 tcgtaaactg aggaggggct ctctgagtat tccaagatat gtagggtagc atcttccacc
3301 gcggatgctg gcgcgcacgt aatcgatatag ttcgtgcgag ggagcgagga ggtcgggacc
3361 gaggttgcta cgggcgggct gctctgctcg gaagactatc tgctgaaga tggcatgtga
3421 gttggatgat atggttggac gctggaagac gttgaagctg gcgtctgtga gacctaccgc
3481 gtcacgcacg aaggaggcgt aggagtcgcg cagcttggtg accagctcgg cggtgacctg
3541 cacgtctagg gcgcagtagt ccagggtttc cttgatgatg tcatacttat cctgcttccc
3601 tttttttcca cagctcgcgg ttgaggacaa actcttcgcg gtctttccag tactcttggg
3661 tcggaaaccc gtcggcctcc gaacggtaag agcctagcat gtagaactgg ttgacggcct
3721 ggtaggcgca gcatcccttt tctacgggta gcgcgtatgc ctgcgcggcc ttccggagcg
3781 aggtgtgggt gagcgcaaaag gtgtccctga ccatgacttt gaggtactgg tatttgaagt
3841 cagtgtcgtc gcatccgccc tgctcccag aagcaaaagt ccgtgcgctt tttggaacgc
3901 ggatttgcca gggcggaagg tgacatcggt gaagagtatt ctttcccgcg cgaggcataa
3961 agttgctgtg gatgcggaag ggtcccggca cctcggaacg gttgttaatt acctgggagg
4021 cgagcacgat ctcgtaaaag ccgttgatgt tgtggcccac aatgtaaagt tccaagaagc
4081 gcgggatgcc cttgatggaa ggcaattttt aagtttcctc gtaggtgagc tcttcagggg
4141 agctgagccc gtgctctgaa agggccagat ctgcaagatg agggttggaa gcgacgaatg
4201 agctccacag gtcacgggcc attagcattt gcagggtggtc gcgaaaggta cctaaactgg
4261 cgacctatgg cctatttttt cttgggggtg atgcagtaga aggtaagcgg gtctgttccc
4321 cagcggtccc atccaagggt cgcggttagg tctcgcgcgg cagtcactag aggctcatct
4381 ccgccgaact tcatgaccag catgaagggc acgagctgct tcccaaaggc ccccatccaa
4441 gtataggtct ctacatcgta ggtgacaaaag agacgctcgg tcgcaggatg cgagccgatc
4501 gggagaact ggatctcccg ccaccaattg gaggagtggc tattgatgtg gtgaaagtaa
4561 gaagtccctg cgacgggccc aacactcgtg cttggctttt gtaaaaactg gcgcagtact
4621 ggcagcgggtg cacgggctgt acatcctgca cgaggttgac ctgacgaccg cgcacaagga
4681 agcagagtgg gaatttgagc cctccgctcg ggcgggtttg gctggtggtc ttctacttcg
4741 gctgcttgte cttgacgctc tggtgctcgg aggggagtta cgggtgatcg gaccaccag
4801 ccgcgcgagc ccaaagtcca gatgtccgcg cgcggcggtc ggagcttgat gacaacatcg
4861 cgcagatggg agctgtccat ggtctggagc tcccgcgcgg tcaggtcagg cgggagctcc
4921 tgcaggttta cctcgcatag acgggtcagg gcggcgggct agatccagg gataccta
4981 ttccaggggc tggttgggtg cggcgctgat ggcttgcaag aggcgcgcat cccgcggcgc
5041 gactacggta ccgcgcggcg ggcgggtggc cggcgggggg gtccttggat gatgcacta
5101 aaagcgggtg cgcgggcgag ccccgaggag gtaggggggg ctccggaccc gccgggagga
5161 gggggcaggg gcacgtcggc gccgcgcgcg ggcaggagct ggtgctgcgc gcgtaggttg
5221 ctggcgaacg cgacgacgcg gcggttgatc tctgaatct ggccgctctg cgtgaagacg
5281 acgggcccgg tgagcttgaa aacctgaaag agagttcgac agaataaatt tcggtgtcgt
5341 tgacggcggc ctggcgcaaa atctcctgca cgtctcctga gttgtcttga taggcgatct
5401 cggccatgaa ctgctcgatc tcttctcctt ggagatctcc gcgtccggct cgtccacgg
5461 tggcggcgag gtcgttggaa atgcgggcca tgagctgcga gaaggcggtt aggcctccct
5521 cgttccagac gcggctgtag accaccgccc ccttcgggca tcgcgggcgc gcatgaccac
5581 ctgcgcgaga ttgagctcca cgtgcggggc gaagacggcg tagtttcgca ggcgtgaaa
5641 gaggtagttg aggggtggtg cgggtgtgtc tgccacgaag aagtacataa cccagcgtcg
5701 caacgtggat tcgttgatat cccccaaggc ctcaaggcgc tccatggcct cgtagaagtc
5761 cacggcgaag ttgaaaaact gggagtgtgc cgccgacacg gttaaactcc cctccagaag
5821 acggatgagc tcggcgacag tgtcgcgcac ctgcgctca aaggctacag gggcctcttc
5881 ttcttcttca atctcctctt ccataagggc ctcccctctt tcttctctg gcggcggtgg
5941 gggagggggg acacggcggc gacgacggcg caacgggagg cggtcgacaa agcgtcgat
6001 catctccccg ccggcgacgg cgcattggtc cggtagcggc gcggcggttc tcggcggggg
6061 cgcagttgga agacgccgcc cgtcatgtcc cggttatggg ttggcggggg ggtgccatg
6121 cggcagggat acggcgctaa cgtatgcatt caacaattgt tgtgtaggta ctccgcgcgc
6181 gagggacctg agcagatccg catcgaccgg atcggaacac ctctcgagaa aggcgtctaa
6241 ccagtcacag tcgcaaggta ggctgagcac cgtggcgggc ggcaggcggg cggcggtcgg
6301 ggttggttct ggcgagggtg ctgctgatga tgtaattaaa gtaggcggtc ttgagacggc
6361 ggatggtcga cagaagcacc atgtccttgg gtccggcctg ctgaatgcgc aggcggtcgg
6421 ccatgccccca ggcttcgttt tgacatcggc gcaggctctt gtagtagtct tgcagagcc
6481 tttctaccgg cacttcttct tctccttcc cttgtcctgc atctcttgca tctatcgctg
6541 cggcgggcgg gcggtttggc cgtaggtggc gccctcttcc tcccatgcgt gtgaccccca
6601 agccctcat cggctgaagc agggctaggt cggcgacaac gcgctcggct aatatggcct
6661 gctgcacctg cgtgagggta gactggaagt catccatgct cacaagcggg ttgtatgcgc
```

```
24721 gtagcctgat tggggagttt acccagccgc cccctgctag ttgagcggga caggggaccc
24781 tgtgtttctca ctgtgatttg caactgtcct aaccctggat tacatcaaga tcctctagtt
24841 aattcagtac ccggggactt acttacctta accctttaac taaataaaaa aataaataaa
24901 agcatcactt acttaaaatc agttagcaaa tttctgtcca gtttattcag cagcacctcc
24961 cttgccctcc ctcccagctc tggatttgca gtttcctcct ggctgcaaac tttctccaca
25021 atactaaata ggaatgtcag tttcctcctg ttctgttcca tccgcacca ctatcttcat
25081 gttgttgtag atgaagcgcg caagaccgtc tgaagatacc ttccaacccc gtgtatccat
25141 atgacacgga aaccggatcc tccaactgtt gctttttctt actcctccct ttgtatcccc
25201 caatgggttt caagagagtc cccctggggt actctctttg cgcctatccg aacctctagt
25261 tacctccaat ggcattgttg cgctcaaaat gggaacggc ctctctctgg acgaggccgg
25321 caaccttacc tcccaaatg taaccactgt gagccacct ctacaaaaaa ccaaagtcaa
25381 acaataaacc tggaaatata tgcaccctc acagttacct cagaagccct aactgtggct
25441 gccgcgcac ctctaattgt cgcgggcaac acactcacca tgcaatcaca ggccccgct
25501 aaccgtgcac gactccaaac ttaagcattg ccaccaagg caccctcca cagtgtcaga
25561 aggaaagcta gccctgcaaa catcaggccc cctccaccac caccgatagc agtaccctta
25621 ctatcactgc ctccaccccc tctaactac tgccactggt agcttgggca ttgacttgaa
25681 agagccatt tatacaacaa aataggaaaa ctaggactaa agtacggggc tctttgtcat
25741 gtaacagacg acctaaacac tttgaccgta gcaactggtc caggtgtgac tattaataat
25801 acttccttgc aaactaaagt tactggagcc ttgggttttg attcacaagg caatatgcaa
25861 cttaatgtag caggaggact aaggattgat tctcaaaaca gacgccttat acttgatgtt
25921 agttatccgt ttgatgtcga aaaccaacta aatctaagac taggacaggg ccttctttt
25981 tattaactc agcccacaac ttggatatta actaacaaca aaggccttta cttgtttaca
26041 gcttcaaaac attccaaaag cttgagggtta acctaaagcac tgccaagggg ttgatgtttg
26101 acgtacagc catagccatt aatgcaggag atgggcttga atttggttca cctaattgcaa
26161 ccaaacaaca aactccccta ccaaaacaaa aattaggcca tggcctagaa tttgattcaa
26221 acaaggctat ggttcctaaa ctagggaactg gccttagttt tgacagcaca ggtgccatta
26281 cagtaaggaa acaaaataat gataagctaa ctttgtggac cacaccagct ccatctccta
26341 actgtagaac taaatgcaga gaaagatgct aaactcactt tggctttaac aaaatgtggc
26401 agtcaaatac ttgctacagt ttctagtttt ggctgttaaa ggcagtttgg ctccaatctc
26461 tggaacagtt acaagtgtc atcttattat aagatttgac agaaatggag tgctactaaa
26521 caattccttc ctggaccag aatattggaa ctttagaaat ggagatctta ctgaaggcac
26581 agcctataca aacgtgttg gatttatgcc taacctatca gcttatacca aaatctcaca
26641 aggtaaaact aagccaaaag taaactattg tcagtctaag ttttaactta acggagaaca
26701 aaactaaacc tgtaacacta accattacac taaacggtac acaaggaaac aggagacaca
26761 actccaagtg catactctat gtctattttc tatgggactg gtctggccac aactaattaa
26821 tgaaatattt gccacatcct cttacacttt tcatacattg cccaagaata aagaatcggt
26881 tggtgttatg tttcaacgtg tttatttttc taattgcaga aaatttcaag tcatttttct
26941 attcagtagt atagccccac caccacatag cttatacaga tcaccgtacc ttaatcaaac
27001 tcacagaacc ctagtattca acctgccacc tccctcccaa cacacagagt acacagtcct
27061 ttcctccccg gcctggcctt aaaaagcaat catatcatgg gtaacagaca tattcttagg
27121 tggtatattc cacacggttt ccttgtcgag ccaaacgctc atcagtgata ttaataaact
27181 cccggggcca gtcacttaa gttcatgtcg ctgtccagct gctgagccac aggctgctgt
27241 ccaacttgcg gttgtttaac gggcggcgaa ggagaagtcc acgcctacat gggggttagg
27301 agtcataatc gtgcatacag ataggcggt ggtgctgcag cagcgcgcg aataaactgt
27361 gccgcgcgcg ctccgtcctg caggaatata acatggcagt ggtctcctca gcgatgattc
27421 gcaccgcccg cagcataagg cgcttgtcc tccgggcaca gcagcgacc ctgatctcac
27481 ttaaatcagc acagtaactg cagcacagca ccacaatatt gttcaaaata cccacagtgc
27541 aaggcgctgt atccaaagct catggcgggg accacagaac ccacgtggcc atcataccac
27601 aagcgcaggt agattaagt ggcaccctc ataaacacgc tggacataaa cattaccttc
27661 ttttggcatg ttgtaattca ccacctccg gtaccatata aacctctgat taaacatggc
27721 gccatccacc accatcctaa accagcctag gccaaacctg cccgcgggcc ctanactaca
27781 ctgcagggaa ccgggactgg aacaatgaca gtggagagcc caggactcgt aaccatggat
27841 catcatgtc gtcatgatat caatgttggc acaacacagg cacacgtgca tacacttctt
27901 caggattaca agctcctccc gcgttagaac catatcccag ggaacaaccc attcctgaat
27961 cagcgtaaat cccacactgc aggaagacc tcgcacgtaa ctacgttgt gcatgtcaa
28021 agtgttacat tcgggcagca gcgatgatc ctccagtatg gtagecgggg tttctgtctc
28081 aaaaggaggt agacgatccc tactgtacgg agtgcgccga gacaaccgag atcgtgttgg
28141 tcgtagtcgt catgccaaat ggaacgcgg acgtagtcat atttctgaa gcaaaaccaa
28201 ggtgcggggc tgacaaacag atctgcgtct ccggtctcgc cgttagatc gctctgtgta
28261 gtagttgtag tatatccact ctctcaaagc atccaggcgc cccctgggt tcgggttcta
```

JOURNAL Submitted (18-AUG-2003) 454 Life Sciences, 20 Commercial Street,
Branford, CT 06405, USA

FEATURES Location/Qualifiers

source 1..1062

/organism="Expression vector pAdEasy-1"

/mol_type="other DNA"

/db_xref="taxon:243021"

/clone="Stratagene catalog number 240005"

/note="contig 2; differs from pAdEasy-1 sequence from
Stratagene; sequenced by new method"

BASE COUNT 282 a 253 c 236 g 290 t 1 others

ORIGIN

```
1 aaattaaaat gaatgtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag
61 ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtcttattt cgttcatcca
121 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccattctggcc
181 ccagtgtctg aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa
241 accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc
301 agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca
361 acgttgttgc cattgctgca ggcacgtgg tgtcacgctc gtcgtttggt atggcttcat
421 tcagctccgg ttccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaagc
481 ggtagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tggtatcact
541 catggttatg gcagcactgc ataattctct tactgtcatg ccattccgtaa gatgcttttc
601 ttgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt
661 gctcttgccc ggcgtcaaca cgggataata ccgcgccaca tagcaagaaa ctttaaaagt
721 tagctcatca ttaggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctggt
781 gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt
841 caccagcgtt tctgggtgag caaaaacagg aaggcaaaaat gccgcaaaaa agggaataag
901 ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta
961 tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat
1021 aggggttccg cgcacatttc cccgaaagtg ccacctgtcn ag
```

//

LOCUS AY370909S3 30091 bp DNA linear SYN 21-AUG-2003

DEFINITION Expression vector pAdEasy-1, contig 3.

ACCESSION AY370911

VERSION AY370911.1 GI:34014919

KEYWORDS .

SEGMENT 3 of 3

SOURCE Expression vector pAdEasy-1

ORGANISM Expression vector pAdEasy-1
artificial sequences; vectors.

REFERENCE 1 (bases 1 to 30091)

AUTHORS Sarkis,G., Costa,G., Leamon,J., Maithreyan,S., Berka,J., Du,L.,
Fierro,J., McDade,K., Puc,B., Roth,G.T., Gomes,X., Altman,W.,
Charumilind,J., Chen,Y.-J., Chen,Z., de Winter,A., Dewell,S.,
Drake,J., Forte,R., He,W., Helgesen,S., Jannotti,M.L., Jarvie,T.,
Jirage,K., Kelch,K., Kim,J.-B., Kukanski,K., Lanza,J., Lee,W.,
Lefkowitz,S., Lu,H., Makhijani,V., Margulies,M., Nobile,J.,
Norton,W., Reifler,M., Rodgers,G., Ronan,M., Simpson,J.,
Tartaro,K., Verma,S., Zimmerman,Z., Dacey,P., Begley,R. and
Lohman,K.

TITLE Sequence Analysis of the pAdEasy-1 Recombinant Adenoviral Construct
Using the 454 Life Sciences Sequencing-by-Synthesis Method

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 30091)

AUTHORS Lohman,K.

TITLE Direct Submission

JOURNAL Submitted (18-AUG-2003) 454 Life Sciences, 20 Commercial Street,
Branford, CT 06405, USA

FEATURES Location/Qualifiers

source 1..30091

/organism="Expression vector pAdEasy-1"

/mol_type="other DNA"

/db_xref="taxon:243021"

/clone="Stratagene catalog number 240005"

/note="contig 3; differs from pAdEasy-1 sequence from
Stratagene; sequenced by new method"

BASE COUNT 6985 a 8742 c 8241 g 6122 t 1 others

ORIGIN

```
1 caggtaggaa gaagtagtat aaggtggggt cttatgtagt tttgtatctt gttttgcagc
61 agccgccgcc gccatgagca ccaactcggt tgatggaagc attgtgagct catatttgac
121 aacgcgcattg ccccatggg cgggggtggc gtcagaatgt gatgggctcc agcattgatg
181 gtgcgcccggt cctgcccgca aactctacta ccttgacctg cgagaccgtg tctggaacgc
241 cgttggagac tgcagcctcc gccgccgctt cagccgctgc agccaccgcc cgcgggattg
301 tgactgactt tgctttcctg agcccgttg caagcagtca gcagcttccc gttcatccgc
361 ccgcgatgac aagttgacgg ctcttttggc acaattggat tctttgacct gggaaacttaa
421 tgtcgtttcg tcagcagctg ttggatctgc gccagcaggt ttctgccctg aaggcttctt
481 cccctcccaa tgcggtttta aacaataaaa taaaaaacca agactctggt tggatatttg
541 atcaagcaag tgtcttgctg tctttattta ggggttttgc gcgcgcggta ggcccgggac
601 cagcgggtctc ggtcgttgag ggtcctgtgt tatttttcca ggacgtggta aaggtgactc
661 tggatgttca gatacatggg cataagcccg tctctggggg ggaggtagca ccactgcaga
721 gcttcatgct ggccggggtg tgtttagat gatccagtcg tagcaggagc ggctgggctg
781 ggtgcctaaa aatgtcttcc agtagcaagc tgattgccag gggcaggccc ttggtgtaag
841 tgtttataaa gcggttaagc tgggatgggt gcatacgtgg ggatatgaga tgcattcttg
901 actgttattt ttaggtttgg ctatgttccc agccatatcc ctccggggat tcatgtttgt
961 cagaaccacc agcacagtgt atccggtgca cttgggaaat ttgtcatgta gcttagaagg
1021 aaatgcgtgg aagaacttgg agacgccctt gtgacctcca agattttcca tgcattcgtc
1081 cataatgatg gcaatggggc cagggcgggc ggccgtggcg aagatatctc tgggataact
1141 aacgtcatag ttgttgttcc aggatgagat cgtcataggc catttttaca aagcgcgggc
1201 gggagggtgc cagactgcgg tataatggtt ccatccggcc caggggctga gttaccctca
1261 cagatttgca tttcccacgc tttgagttca gatgggggga tcatgtctac ctgcggggcg
1321 atgaagaaaa cggtttccgg ggtaggggag atcagctggg aagaaagcag gttcctgagc
1381 agctgcgact taccgcagcc ggtgggcccg taaatcacac ctattaccgg cctgcaactg
1441 gtagttaaga gagctgcagc tgccgtcatc cctgaggcag gggggccact tcgtaagca
1501 tgtccctgac tcgcatgttt tccctgacca aatccgccag aaggcgctcg ccgccagcg
1561 atagcagttc ttgcaaggaa gcaaagtttt tcaacgggtt gagaccgtcc gccgtaggca
1621 tgcttttgag cgtttgacca agcagttcca ggccgtccca cagctcggtc acctgctcta
1681 cggcatctcg atccagcata tctcctcggt tcgcgggttg gggcggtttt cgtgtacgg
1741 cagtagtcgg tgctcgtcca gacgggccag ggtcatgtct ttccacgggc gccagggtcc
1801 tcgtcagcgt agtctgggtc acggtgaagg ggtgcgtccc gggctgcgcg ctggccaggg
1861 tgccgttgag gctggtcctg ctggtgctga agcgtgcgcg gtcttcgccc tgcgcgtcgg
1921 ccaggtagca tttgaccgat ggtgtcatag tccagcccct ccgcggcgtg gcccttggcg
1981 cgcagcttgc ccttgaggga ggccgcgcac gaggggcagt gcagactttt gagggcgtag
2041 agcttgggcg cgagaaatac cgattccggg gaggtaggca tccgcgcgcg caggccccgc
2101 agacggtctc gcattccacg agccaggtga gctctggccg ttcgggttca aaaaccaggc
2161 tttcccccatt tgctttttga tgcgtttctt acctctggtt tccatgagcc ggtgtccacg
2221 ctccgtgacg aaaaggctgt ccgtgtcccc gtatacagac ttgagaggcc tgtcctcgag
2281 cgggtgtccg cggtcctcct cgtatagaaa ctccgaccac tctgagacaa aggtcgcgt
2341 ccaggccagc acgaaggagg ctaagggtgg aggggtaggc ggtcgttgtc cactaggggg
2401 tccactcgct ccagggtgtg aagacacatg tcgccctctt cggcatcaag gaaggtgatt
2461 ggtttgtagg ttagggccac gtgaccgggt gttcctggaa ggggggctaa gtaaaagggg
2521 gtggggggcg cgttcgtcct cactctcttc cgcctcgtg tctgcgaggg ccagctgttg
2581 ggggtgagta tccctctgaa aagcgggcat gacttctgcg ctaagattgt cagtttacca
2641 aaaacagagg aggatttgat attcacctgg cccgcggtga tgcctttgag ggtggccgca
2701 tccatctggt caagaaaaga acaatctttt gttgtcaagc ttggtggcaa acgaccgta
2761 gaggcggttg gacagcaact tggcgatgga gcggcagggt ttggtttttg ttcgcgatcg
2821 gcgcgtcctt tggccgcgat gtttagctgc acgtattcgc gcgcaacgca ccgccattcg
2881 ggaaagacgg tggtcgcgtc gtcgggcacc aggtgcacgc gccaacgcgc gttgtgcagg
2941 gtgacaaggt caacgctggt ggctacctct ccgcgtaggc gctcgttggt ccagcagagg
3001 cggccgccct tgcgcgagca gaatggcggg tagggggtct agctgcgtct cgtgccgggg
3061 ggtctgcgtc cacggtaaaag accccgggca gcaggcgcg ctcgaagtag tctatcttgc
```

```
6721 ccgtgttgat ggtgtaagtg cagttggcca taacggacca gttaacggtc tggtagcccg
6781 gctgcgagag ctcggtgtac ctgagacgcg agtaagccct cgagtcaa atacgtagtcgt
6841 tgcaagtccg caccaggtac tggatcccca ccaaaagtgc ggccggcggt ggccggttagga
6901 ggggccaggc gtaggggtgg cgggggctgc cgggggcgga gatcttccaa cataaggcga
6961 tgatatccgt agatgtacct ggacatccag gtgatgccgg cggcgggtgt ggaggcgcg
7021 ggaaagtgcg ggacgcgggt ccagatgttg cgcagcaggc aaaagtagct ccatggtcgg
7081 gagctctctg ccggtcaggc gcgcgcaatc gttgacgctc tagcgtgcaa aaggagagcc
7141 gttaagcggg cactcttccg tggctctggt gataaattcg caagggtatc atggcggacg
7201 accgggggtc gagccccgta tccggccgct cgccgtgacg catgcggtta ccgcccgcgt
7261 gtcgaaacca ggtgtgcgac gtcagacaag cgggggagtg ctctcttttg gcttcttcc
7321 aggcgcggcg gctgctgcgc ttagcttttt ggccactggc cgcgcgcagc gtaagcgggt
7381 aggctggaaa gcgaaagcat taagtggctc gctccctgta gccggagggt tattttccaa
7441 gggttgagtc gcgggacccc cggttcgagt ctcgacccg cgggactgcg gcgaacgggg
7501 gtttgccctc ccgtcatgca agaccccgtc tgcaaattcc tccggaaaca gggacgagcc
7561 ccttttttgc cttttccag atgcatccgg tgcgaggca gatgcgcccc cctccctcag
7621 cagcggcaag agcaagagca gcggcagaca tgcagggcac cctcccctcc ctctaccgc
7681 gtcaggaggg gcgacatccg cggttgacgc ggcagcagat ggtgattacg aacccccgcc
7741 ggcgcggggc ccggcactac ctggacttgg aggaggcgcg agggcctggc gcggctagga
7801 gcgcctctc ctgagcggcc caccgaagg atgcagctga agcgtgatac gcgtgaggcg
7861 tacgtgccgc ggcagaacct gtttcgcgac cgcgagggag aggagcccga ggagatgcgg
7921 gatcgaaagt tccacgcagg gcgcgagctg cggcatggcc tgaatcgcg gcggttgctg
7981 cgcgaggagg actttgagcc cgacgcgcga accgggatta gtcccgcgcg gcacacgtg
8041 gcggccgcgc acctggtaac cgcatacagc cagacgggtg accaggagat taactttaca
8101 aaaagcattt aaacaaccac gtgcgtacgc ttgtggcgcg cgaggagggt gctataggac
8161 tgatgcatct gtgggacttt gtaagcgcgc tggagcaaaa cccaaataag caagccgctc
8221 atggcgcagc tgttccttat agtgcagcac agcagggaca acgaggcatt cagggatgcg
8281 ctgctaatac tagtagagcc cgaggcgccg tggctgctcg atttgataaa catctgcag
8341 agcatatagg tgcaggagcg cagcttgagc ctggctgaca aggtggcgcc catcaactat
8401 tccatgctta gcctgggcaa gttttacgcc cgcaagatat accatacccc ttacgttccc
8461 atagacaagg aggtaaagat cgagggggtc tacatgcgca tggcgctgaa tggtgcttac
8521 cttgagcgac gacctgggcg tttatcgcaa cgagcgcac cacaaggccg tgagcgtgag
8581 ccggcggcgc gagctcagcg accgcgagct gatgcacagc ctgcaaagg ccctggctgg
8641 cacgggcagc ggcgatagag aggcgagtc ctactttgac gcgggcgctg acctgcgctg
8701 ggccccaaac cgacgcgccc tggaggcagc tggggccgga cctgggctgg cggtgccacc
8761 cgcgcgcgct ggcaacgtcg gcggcgtgga ggaatatgac gaggacgatg agtacgagcc
8821 agaggacggc gagtactaag cggatgatgt tctgatcaga tgatgcaaga cgcaacggac
8881 ccggcgggtg gggcggcgct gcagagccag ccgtccggcc ttaactccac ggacgactgg
8941 cgccaggta tggaccgcat catgtcgtcg actgcgcgca atcctgacgc gttccggcag
9001 cagccgcagg ccaaccggct ctccgcaatt ctggaagcgg tggccccggc gcgcgcaaac
9061 cccacgcacg agaaggtgct ggcgatcgta aacgcgctgg ccgaaaacag ggccatccgg
9121 cccgacgagg ccggcctggt ctacgacgcg ctgcttcagc gcgtggctcg ttacaacagc
9181 ggcaacgtgc agaccaacct ggaacggctg gtgggggatg tgcgcgagcg cgtggcgagc
9241 cgtgagcgcg cgcagcagca gggcaacctg ggctccatgg ttgcaactaa cgccttctcg
9301 agtacacagc ccgccaacgt gccgcgggga caggaggact acaccaactt tgtgagcgca
9361 ctgcggctaa tggtgactga gacaccgcaa agtgagggtg accagtctgg gccagactat
9421 ttttccagac cagtagacaa ggcctgcaga ccgtaaacct gagccaggct ttcaaaaact
9481 tgcaggggct ggtgggggtt ggccgggtcc cacaggcgac cgcgcgaccg tgtctagctt
9541 gctgacgccc aactcgcgcc tgttgctgct gctaatagcg cccttcacgg acagtggcag
9601 cgtgtcccg gacacatacc taggtcactt gctgacactg taccgcgagg ccataggtca
9661 ggcgcatgtg gacgagcata ctttccagga gattacaagt gtcagccgcg cgtggggca
9721 ggaggacacg ggcagcctgg aggcacacct aaactacctg ctgaccaacc ggcggcagaa
9781 gatccccctg ttgcacagtt taacacgcga ggaggagcgc attttgcgt acgtgcagca
9841 gagcgtgagc cttaacctga tgcgcgacgg ggtaacgccc agcgtggcgc tggacatgac
9901 cgcgcgcaac atggaaccgg gcatgtatgc ctcaaaccgg ccgtttatca accgccta at
9961 ggactacttg catcgcgcgg ccgccgtgaa ccccgagtat ttcaccaatg ccatcttgaa
10021 cccgcactgg ctaccgcccc ctggtttcta caccggggga ttcgagggtg ccgagggtaa
10081 cgatggattc ctctgggacg acatagacga cagcgtgttt tccccgcaac cgcagaccct
10141 gctagagttg caacagcgcg agcaggcaga ggccgcgctg cgaaaggaaa gacttccgca
10201 ggccaagcag cttgtccgat ctaggcgctg cggccccgcg gtcagatgct agtagcccat
10261 ttccaagctt gatagggtct cttaccagca ctgcaccac ccgcccgcgc ctgctgggcg
```

```
10321 aggaggagta cctaaacaac tcgctgctgc agccgcagcg cgaaaaaacc tagcctccgg
10381 catttcccaa ccaacgggat agagagccta gtggacaaga tgagtagatg gaagacgtac
10441 gcgcaggagc acagggacgt gccaggcccc ccgcccgcgc acccgctcgtc aaaggcacga
10501 ccgtcagcgg ggtctggtgt gggaggacga tgactcggca gacgacagca gcgtcctgga
10561 tttgggaggg agtggcaacc cgtttgcgca ccttcgcccc aggcctgggg aggaatagtt
10621 ttaaaaaaaa aaaaatagat gcaaaaataa aaactacacc aaggccatgg caccgagcgt
10681 tggttttctt gtattcccct tagtatgcgg cgcgcggcga tgtatgagga aggtcctcct
10741 cctcctaag agagtgtggt gagcgcggcg ccagtggcgg cggcggctgg gttctccctt
10801 cgatgctccc ctggaccgcg cgtttgtgcc tccgcggtac ctgcggccta ccggggggag
10861 aaacagcatc cgttactctg agttggcacc cctattcgac accaccctgt tgtacctggt
10921 ggacaacaag tcaacggatg tggcatccct gaactaccag aacgaccaca gcaactttct
10981 gaccacggtc attcaaaaca atgactacag cccgggggag gcaagcacac agaccatcaa
11041 tcttgacgac cggtcgcaca tggggcggcg acctgaaaac catcctgcat accaacatgc
11101 caaatgtgaa cgagttcatt tttaccaata agtttaaggc ggcggtgat ggtgtcgcgc
11161 ttgcctacta aggacaatca ggtggagctg aaatacagat ggggtgagtt cacgctgcc
11221 gagggcaact actccgagac catgaccata gaccttatga acaacgcgat cgtggagcac
11281 tacttgaaag tgggcagaca gaacggggtt gctggaaagc gacatcgggg taaaggtttg
11341 acaccgcgaa cttcagactg gggtttgacc ccgtcactgg tcttgctcat cctggggtat
11401 atacaaacga agccttccat ccagacatca ttttgctgcc aggatgcggg gtggacttca
11461 cccacagccg cctgagcaac ttgttgggca tccgcaagcg gcaacccttc caggagggct
11521 ttaggatcac ctacgatgat ctggagggtg gtaacattcc cgcactgttg gatgtggacg
11581 cctaccaggc gagcttgaaa gatgacaccg aacaggggcg ggggtggcga ggcggcagca
11641 acagcagtgg cagcggcgcg gaagagaact ccaacgcggc agccgcggca atgcagccgg
11701 tggaggacat gaacgatcat gccattcgcg gcgacacctt tgccacacgg gctgaggaga
11761 agcgcgctga ggccgaagca gcggccgaag ctgccgcccc cgctgcgca acccgaggtc
11821 gagaacctc agaagaaacc ggtgatcaaa cccctgacag aggacagcaa gaaacgcagt
11881 tacaaccta taagcaatga cagcaccttc acccagtacc gcagctggtg ccttgcatat
11941 aactacggcg accctcagac cggaatccgc tcatggacce tgctttgcac tctgacgta
12001 acctgcggct cggagcaggt ctactggtcg ttgccagaca tgatgcaaga ccccgtagcc
12061 ttccgctcca cgcgccagat cagcaacttt ccggtggtgg gcgcgagct gttgcccggtg
12121 cactccaaga gcttctacaa cgaccaggcc gtctactccc aactcatccg ccagtttacc
12181 tctctgacct acgtgttcaa tcgctttccc gagaaccaga ttttggcgcg cccgcccagc
12241 ccccaccatc accaccgtca gtgaaaacgt tctgtctctc acagatcacg ggacgctacc
12301 gctgcgcaac agcatcggag gagtccagcg agtgaccatt actgacgcca gacgcgcac
12361 ctgcccctac gtttacaagg ccctgggcat agtctcgccg cgcgtcctat cgagccgcac
12421 ttttgagcaa gcatgtccat ccttatatcg cccagcaata acacaggctg gggcctgcgc
12481 ttcccaagca agatgttttg cggggccaag aagcgtcccg accaacaccc agtgcgcggtg
12541 cgcgggcact accgcgcgcc ctggggcgcg caaaaacgcg gccgcaactg gcgcaccacc
12601 gtcgatgacg ccatcgacgc ggtggtggag gaggcgcgca actacacgce caccgcgcgc
12661 accagtgtcc acagtggacg cggccattca gaccgtggtg cgcggagccc ggcgctatgc
12721 taaaatgaag agacggcgga ggcgcgtagc acgtcgccac cgcgcgcgac ccggcactgc
12781 cgcccaacgc gcggcgcgcg cctgcttaa ccgcgcagct cgcaccggcc gacggggcgc
12841 catgcggggc gtcgaaggc tggcgcgggg tattgtcact gtgcccccca ggtccaggcg
12901 acgagcgccc gccgagcag ccgcttagcg cgtgctatg actcagggtc gcaggggcaa
12961 cgtgtattgg gtgcgcgact cggttagcgg cctgcgcgtg cccgtgcgca ccccgcccc
13021 cgccgcaact agattgcaag aaaaaactaa cttagactcg tactgttgta tgtatccagc
13081 ggcgggcgcg cgcaacgaag ctatgtccaa gcgcaaaatc aaagaagaga tgctccaggt
13141 catcgcgccg gagatctatg gcccccgaa gaaggaagag caggattaca agccccgaaa
13201 gctaaagcgg gtcaaaaaga aaaagaaaag aatgatgatg atgaacttga cgacgaggtg
13261 gaactgctgc acgctaccgc gccaggcga cgggtacagt ggaaaggtcg acgcgtaaaa
13321 cgtgttttgc gaccggcac caccgtagtc tttacgcccg gtgagcgctc caccggccac
13381 ctacaagcgc gtgtatgatg aggtgtacgg cgacgaggac ctgcttgagc acggccaacg
13441 agcgctcgg ggagtttgcc tacggaaagc ggcataagga catgctggcg ttgccgctgg
13501 acgagggcaa cccaacacct agcctaaagc ccgtaacact gcagcaggtg ctgcccgcgc
13561 ttgcaccgtc cgaagaaaag cgcggcctaa agcgcgagtc tggtgacttg gcaccaccg
13621 tgcagctgat ggtacccaag cgccagcgac tggagatgt cttggaaaaa atgaccgtgg
13681 aacctgggct ggagcccag gtccgcgtgc ggccaatcaa gcaggtggcg ccgggactgg
13741 gcgtgcagac cgtggacgtt cagataacca ctaccagtag caccagtatt gccaccgcca
13801 cagagggcat ggagacacaa acgtccccgg ttgcctcagc ggtggcggat gccgcgggtg
13861 aggcggctgc tcggccgcg tccaagacct ctacggaggt gcaaacggac ccgtggatgt
```



```
13921 ttccgcgttttc cagcccccccg gccgcccgcg ccgttcgagg aagtacggcg ccgccagcgc
13981 gctactgccc gaatatgccc tacatccttc cattgcgcct acccccggct atcgtggcta
14041 cacctaccgc ccagagaagac gagcaactac ccgacgccga accaccactg gaaccgcgcg
14101 ccgccgtcgc cgtcgccagc ccgtgctggc ccgattttcc gtgcgcaggg tggctcgcga
14161 aggaggcagg accctggtgc tgccaacagc gcgctaccac ccagcatcg tttaaaagcc
14221 ggtctttgtg gttcttgcag atatggccct cactgcccgc ctccgtttcc cgtgcccggg
14281 attccgagga agaatgcacc gtaggagggg catggccggc caggcctga cgggcggcat
14341 gcgtcgtgcg caccaccggc ggccgcccgc gtcgcacgt cgcgtgcgcg gcggtatcct
14401 gcccctcctt attccactga tcgccgcggc gattggcgcc gtgccggaa ttgcatccgt
14461 ggcccttgag gcgcagagac actgaattaa aaacaagttg catgtggaaa ataacaaaat
14521 aaaaagtact ggactctcac gctcgtttgg tctgttaact atttttaga atggaagaca
14581 tcaactttgc gtctctggcc ccgccgacac ggctcgcgc cgttcatggg aaactggcaa
14641 gatatacgga ccagcaatat gagcgttggc gccttcagct ggggctcgt gtggagcggc
14701 aattaaaaat ttccggttcca ccgttaagaa ctatggcagc aaggcctgga acagcagcac
14761 aggccagatg ctgagggata agttgaaaga agcaaaattt ccaacaaaag gtggtagatg
14821 gcctggcctc tggcattagc ggggtgggtg gacctggcca accaggcagt gcaaaataag
14881 attaacagta agcttgatcc ccgcctccc gtagaggagc ctccaccggc cgtggagaca
14941 gtgtctccag aggggcggtg gcgaaaagcg tccgcgcgcc cgacagggaa gaaactctgg
15001 tgacgcaaat agacgagcct ccctcgtacg aggaggcact aaagcaaggc ctgcccacca
15061 cccgtcccat cgcgcccatg gctaccggag tgctgggcca gcacacaccc gtaacgctgg
15121 acctgccctc ccccgcgcgc acaccagca gaaacctgtg ctgccaggcc cgaccgcctg
15181 tgttgtaacc cgtcctagcc gcgcgtccct gcgcgcgcgc gccagcggtc cgcgatcgtt
15241 gcggcccgta gccagtggca actggcaaaag cactactgaac agcatcgtgg gtgctggggg
15301 tgcaatccct gaagcgcgga cgtgcttct gatagctaac gtgtcgtatg tgtgtcatgt
15361 atgcgtccat gtcgcgcgca gaggagctgc tgagccgcgc cgcgcgcgct ttccaagatg
15421 gctaccctt cgtgatgcc gcagtggctt tacatgcaca tctcgggcca ggacgcctcg
15481 gagtaccctga gcccgggct ggtgcagttt gccgcgcgca ccgagacgta ctcagcctg
15541 aataacaagt ttaagaaacc ccacggtggc gcctacgcac gacgtgacca cagaccggtc
15601 ccagcgtttg acgctgcggt tcatccctgt ggaccgtgag gatactgctg actcgtacaa
15661 ggcgcggttc accctagctg tgggtgataa ccgtgtgctg gacatggctt ccacgtactt
15721 tgacatccgc ggcgtgctgg acaggggccc tacttttaag cccttactct ggcactgctt
15781 acaacgcctt ggctcccaa ggggtgcccc aatccttgcg aatgggatga agctgctact
15841 gctcttgaat taaacctaag aagaagagga cgatgacaac gaagacgaag tagacgagca
15901 agctgagcaa gcaaaaacta cactgatttt ggcaggcgcc ttattctggt ataaatatta
15961 caaaggaggg tattcaaata ggtgtcgaag gtcaaacaac ctaaatatgc cgataaaaca
16021 atttcaacct gaacctcaaa taaggagaat ctacgtggta cgaaacaaga aattaaatca
16081 tgcagctggg aggagtacct aaaaaagcac taccccaatg aaacatggtt acggttcata
16141 tgcaaaaccc acaaatgaaa atggagggca aggcattctt gtaaagcaac aaaatggaaa
16201 gctagaaagt aaagtggaaa tgaatttttt ctaactaact agtaggcagg ccagccgcag
16261 gcaatggtga ttaacttact acctaagtgg tattgtacag tgaagatgta gatataagaa
16321 accccagaca ctcatatttc ttacatgccc actattaagg aaggtaactc acgagaacta
16381 atgggccaac aatctatgcc caacaggcct aattacattg cttttaggga caattttatt
16441 ggtctaattg attacaacag cacgggtaat atgggtgttc tggcgggcca agcatcgcag
16501 ttgaattcgt ttgtagattt gcaagacaag aaacaacaga gctttcatac cagcttttgc
16561 ttgattccat tggtagata tttctatgtg gaatcaggct gttgacagct
16621 atgatccaga tgttagaatt attgaaatac atggaactag aagatgaact taccaaatta
16681 ctgctttcca ctgggaggtg tgattaatac agagactctt accaaggtaa aacctaaaac
16741 aggtcaggaa aataggatgg gaaaagaata gctacagaat tttacttaga taaaataaga
16801 aataagagtt ggaaataatt ttgccatgga aatcaatcta aatgccaaac tgtggagaaa
16861 tttcctgtac tccaacatag cgtctgattt gcccgacaag cttaaagtaca gtccttccaa
16921 cagtaaaaaat ttctgataac ccaaacaacc tacgactaca tgaacaagcg agtgggtggc
16981 cccgggcctc agtggactgc tacattaacc ttggagcacg ctggtccctt gactatatgg
17041 acaacgtcaa ccattttaac caccaccgca atgctggcct gcgctaccgc tcaatgttgc
17101 tgggcaatgg tcgctatgtg cccttcaca tccaggtgcc tcagaagttc tttgccatta
17161 aaaacctcct tctcctgccg ggctcataca cctacgagtg gaacttcagg aaggatgtta
17221 acatggttct gcagagctcc ctaggaaatg acctaagggt tgacggagcc agcattaagt
17281 ttgatagcat ttgcctttac gccaccttcc ttccccatgg ccacacaacac cgcctccacg
17341 cttgaggcca tgcttagaaa cgacaccaac gaccagtcc ttaacgacta tctctccgcc
17401 gccaacatgt ctctacccta taccgcgcaa cgctaccaac gtgccatat ccattccctc
17461 ccgcaactgg gcggctttcc gcggctgggc cttcacgcgc cttaaagacta aggaaacccc
```

17521 atcactgggc tcgggctacg acccttatta cacctactct ggctctatac cctacctaga
17581 tggaaaccttt tacctcaacc acacctttaa gaagggtggc attacctttg actcttctgt
17641 cagctggcct ggcaatgacc gcctgcctta cccccaaccg agtttgaaat taagcgctca
17701 gttgacgggg aggggttaca cgctgcccag tgtaacatga ccaaagactg gttcctggta
17761 caaatgctag ctaactatta acattggcta ccagggtctc tatatcccag agagctacaa
17821 ggaccgcatg tactccttct ttagaaactt accagcccat gagccgtcag gtgggtgatg
17881 ataactaaata acaaggacta ccaacagggt ggcatcctac accaacaaca acaactctgg
17941 atttgtttggc taccttgccc ccaccatgcg cgaaggacag gcctaccctg ctaacttccc
18001 ctatccgctt ataggcaaga ccgcagttga cagcattacc caagaaaagt ttactttgcg
18061 atcgcaccct ttggcgcac ccatctctca gtaactttat gtccatgggc gactcacag
18121 acctgggcca aaaccttctc tacgccaaact ccgcccacgc gctagacatg actttgaggt
18181 ggatcccatg gacgagccca ccttctttta ttgttttgtt tgaagtcttt gacgtggtcc
18241 gtgtgacacc aagcgcgacc cggcgtcat cgaaccgtg tacctgcgca cgcccttctc
18301 ggccggcaac gccacaacat aaagaagcaa gcaacatcaa caacagctgc cgccatgggc
18361 tccagtggagc aggaactgaa agccattgtc aaagatcttg gttggtgggc cattattttt
18421 gggctacctt tgacaagcgc ttctctaggc ttgttttctc cacacaagct cgctgcgcc
18481 atagtcaata cggcgggtcg cgagactggg ggcgtacact ggatggcctt tgccctggaac
18541 ccgcaactca aaacatgcta cctctttgag ccctttggct tttctgacca gcgactcaag
18601 caggtttacc agtttgagta cgagtcactc ctgcgccgta gcgccattgc ttccttcccc
18661 cgaccgctgt ataacgctgg aaaagtccac ccaaagcgta cagggggcca actcggccgc
18721 ctgtggacta ttctgtgca tgtttctcca cgctttgccc aacctggccc caaacctccc
18781 atggatcaca accccaccat gaaccttatt accgggttac ccaactccat gctcaacagt
18841 ccccggttac agcccacctt gcgtcgcaac caggaacagc tctacagctt cctggagcgc
18901 cactcgccct acttcccgag ccacagtgcg cagattagga gcgccacttc tttttgttca
18961 cttgaaaaac aatagtaaaa ataagtact agagacactt tcaataaagg caaattgctt
19021 ttattttgta cactctcggt tgattattta cccccacctt tgccgtctgc gccgtttaa
19081 aatcaaaggg gttctgccc gcctcgctat gcgccactgg cagggacacg ttgcgatact
19141 ggtgtttagt gctccactta aactcaggca caaccatccg cggcagctcg gtgaagtttt
19201 cactccacag gctgcgcacc atcaccacgc cgtttagcag gtcgggcgcc gatattctga
19261 agtcgcagtt ggggcctccg cctgcgcgc gcgagttgcg atacacaggg ttgcagcact
19321 ggaacactat cagcgcgggg tgggtgcacgc tggccagcac gctcttgctg gagatcagat
19381 ccgcttcag gtcctccggt ttgctcaggg cgaacggagt caactttggt agctgccttc
19441 ccaaaaaggg cgcgtgccc ggctttgagt tgcaactcgc ccgtagtggc atcaaaaagg
19501 gaccgtgccc ggtctggggc ttaggataca gcgcctgcat aaaagccttg atctgcttaa
19561 aagccacctg agcctttgcg ccttcagaga agaacatgcc gcaagacttg ccggaaaact
19621 gattggccgg acaggccggc tegtgcacgc agcaecttgc gtcggtgttg gagatctgca
19681 ccacatttcg gccccaccgg ttcttcacga tcttggcctt gctagactgc tcttcagcg
19741 cgcgtgccc gttttcgctc gtcacatcca ttccaatcac gtgctcctta tttatcataa
19801 tgcttcggtg tagacactta agctcgctt cgatctcagc gcagcgggtg agccacaacg
19861 cgcagcccggt gggctcgtga tgcttgtagg tcacctctgc aaacgactgc aggtacgcct
19921 gcaggaatcg ccccatcatc gtcacaaagg tcttggtgct ggtgaaggtc agctgcaacc
19981 cgcggtgctc ctggttcagc caggctcttc atacggccgc cagagcttcc acttggtcag
20041 gcagtagttt gaagttcgcc tttagatcgt tatccacgtg gtacttgctc atcagcgcgc
20101 gcgcagcctc catgcccttc ctcccacgca gacacgatcg gcacactcag cgggttcate
20161 accgtaattt cactttccgc ttctgtgggc tcttctctt cctcttgctg ccgcatacca
20221 cgcgccactg ggtcgtcttc attcagccgc cgcactgtgc gcttacctcc tttgccatgc
20281 ttgattagca ccggtgggtt gctgaaaccc accatttgta gcgccacac ttcttcttcc
20341 ttctcgtctg tccacgatta cctctggtga tggcgggcgc tcgggcttgg gaggaagggc
20401 gcttcttttt ctttcttggg cgcaatggcc aaatccgcgc ccgaggtcga tggccgcggg
20461 ctgggtgtgc gcggcaccag cgcgtcttgt gatgagttt cctcgtcttc ggactcgata
20521 cgcgcctca tctgttttt tgggggcggc ccggggaggg cggcggcgac ggggagcggg
20581 gacgacagct cctccatggt tgggggagcg tcgcgcgcga ccgcgtccgc gctcgggggt
20641 ggtttcgcgc tgctcctctt cccgactggc catttcttc tctataggc agaaaaagaa
20701 tcatggagtc agtcgagaag aaggacagcc taaccgcgcc cctcctgagt tcgccaccac
20761 cgcctccacc gatgcgcga acgcgcctac cacccttccc cgtcgaggca ccccgcttg
20821 aggaggagga agtgattatc gagcaggacc caggttttgt taagcgaaga cgacgaggac
20881 cgctcagtac caacagagga ataaaaagca aagaaccagg acaacgcaga ggcaaacgag
20941 gaacaagtcg ggcgggggga gcgaaaggca tggcgactac ctatagtggt gagacgagct
21001 gctgttgag catctgcag gccagtgcgc cattatctgc gacgcgttgc aagagcgcag
21061 cgatgtgccc ctccgccata gcggtgttca gccttgccca cgaacgccac ctattctcac

```
21121 cgcgcgtacc ccccaaaccg ccaagaaaac ggcacatgcg agcccaaccc gccgcctcaa
21181 cttcctaccc cgtatttgcc gtgccagagg tgcttgccac ctatcacatc tttttccaaa
21241 actgcaagat acccctatcc tgccgtgcc aaccgagccg agcggacaag cagctggcct
21301 tgcggcaggg cgctgtcata cctgatatcg cctcgctcaa cgaagtgcc aaaaatctttg
21361 aggggtgctt gacgcgacga gaagcgcgcg gcaaacgctc tgcaacaagg aaaacaagcg
21421 aatagaaaag tcaactctga gtgttggtgg aactcgaggg tgacaacgcg cgcctagccg
21481 tactaaaacg cagcatcgag gtcacccact ttgcctaccc ggcaactaac ctacccccc
21541 aggtcatgag cacagtcag agtgagctga tcgtgcgcgc tgcgcagccc ctggagaggg
21601 atgcaaattt gcaagaaaca aacaagagga gggcctaccc gcagttggcg acgagcagct
21661 agcgcgctgg cttcaaaccg gcgagcctgc cgacttgag gagcgacgca aactaatgat
21721 ggccgcagtg ctcgttaccg tggagcttga gtgcatgcag cggttctttg ctgaccggga
21781 gatgcagcgc aagctagagg aaacattgca ctacaccttt cgacagggct acgtacgcca
21841 ggccgtgcaag atctccaacg tggagctctg caacctggtc tcctaccttg gaattttgca
21901 cgaaaaccgc cttgggcaaa acgtgcttca ttccacgctc aagggcggag gcgcgcgcg
21961 actacgtccg cgactgcgtt tacttatttc tatgtacac ctggcagacg gccatgggcg
22021 tttggcagca gtgcttgag gagtgcaacc tcaaggagct gcagaaactg ctaaagcaaa
22081 acttgaagga cctatggacg gccttcaacg agcgtccgt ggccgcgcac ctggcgga
22141 tcattttccc cgaacgcctg cttaaaaccc tgcaacaggg tctgccagac ttcaccagtc
22201 aaagcatgtt gcagaacttt aggaacttta tctagagcg ctcaggaatc ttgcccgcca
22261 cctgctgtgc acttcctagc gactttgtgc ccattaagta ccgcgaatgc cctccgcgcg
22321 tttggggcca ctgctacctt ctgcagctag ccaactacct tgcctaccac tctgacataa
22381 tggaagacgt gagcggtgac ggtctactgg agtgtcactg tcgctgcaac ctatgcaccc
22441 cgccaccgct ccctggtttg caattcgag ctgcttaacg aaagtcaa atatcgggtacc
22501 tttgagctgc aggggtccctc cgctgacga aaagtccgcg gctccggggt tgaaactcac
22561 tccggggctg tggacgtcgg cttaccttcg caaatttgta cctgaggact accacgccc
22621 cgagattagg ttctacgaag accaatcccg ccgcctaata gcggagctta ccgcctgctg
22681 cattaccag ggccacattc ttggccaatt gcaagccatc aacaaagccc gccaagagtt
22741 tctgctacga aagggacggg ggggttactt ggaacccag tccggcgagg tccctaacc
22801 aatccccccg ccgcgcgcag ccctatcagc agcagccgcg ggcccttgcc tcccaggat
22861 ggcacccaaa aagaagctgc agctgcgcgc gccacccacc ggacgaggag gaatactggg
22921 acagtcaggc agaggaggtt ttggacgagg aggaggagga catgatggaa gactgggaga
22981 gcctagacga ggaagcttcc gaggtcgaag aggtgtcaga cgaaacaccg tcacctcgg
23041 tcgcattccc ctccgcgcgc cgcccagaa atcggcaacc ggttccagca tggctacaac
23101 ctccgctcct caggcgcgcg cggcactgcc cgttcgcga cccaaccgta gatgggacac
23161 cactggaacc agggccggta agtccaagca gccgcgcgcg ttagcccaag agcaacaaca
23221 gcgccaaggc taccgctcat ggcggcgggc acaagaacgc catagttgct tgcttgcaag
23281 actgtggggg caacatctcc ttccgcgcgc cgctttcttc tctaccatca cggcgtggcc
23341 ttcccccgta acatcctgca ttactaccgt catctctaca gccatactg caccggcggc
23401 cggcggcagc aaagaacagc aacagcagcg gccacacaga agcaaaggcg accggatagc
23461 aagactctga caaagcccaa gaaatccaca gcggcggcag cagcaggagg aggagcgtg
23521 cgtctggcgc ccaacgaacc cgtatcgacc cgcgagctta gaaacaggat tttccactc
23581 tgtatgtcta tatttcaaca gagcaggggc caagaacaag agctgaaaat aaaaaacaag
23641 gtctctgcga tccctccacc cgcagctgcc tgtatcaca aagcgaagat cagcttcggc
23701 gcacgctgga agacgcggag gctctcttca gtaataactg cgcgctgact cttaggact
23761 agtttcgcgc cctttctcaa atttaagcgc gaaaactacg tcatctccag cggccaacc
23821 cggcgcgcag ccacctgctt gttgtcagcg ccattatgag caaggaaatt cccaccgccc
23881 tacatgtgga gttaccagcc acaaatggga cttgcggctg gagctgccc aagactactca
23941 accgaataa actacatgag cgcgggaccc cacatgatat ccgggtcaa cggaataacg
24001 cgccaccgca aaccgaattc tcctggaaca ggcggtatt accaccacac ctcgtaataa
24061 ccttaatccc cgtagtggc cgccctgcc tgggtgacca ggaaagtccc gcctcccacc
24121 actgtggtac ttcccagaga cgcccaggcc gaagttcaga tgactaactc aggggcggca
24181 gcttgccggc ggctttcgtc acaggggtgg ggtcgcccg gcagggtata actcacctga
24241 caatcagagg gcggaggata ttcagactca acgacgagtc ggtgagctcc tcgcttggtc
24301 tccgtccgga cgggacattt cagatcggcg gcgcgcgcgc cctcttcatt cagcctcgt
24361 caggcaatcc taactctgca gacctcgtcc tctgagccgc gctctggagg cattggaact
24421 ctgcaattta ttgaggagtt tgtgccatcg gtctacttta accccttct cgggacctcc
24481 cgccactat ccggatcaat ttattcctaa ctttgacgcg gtaaaggact cggcgagcgg
24541 ctacgactga atgttaagt gagaggcaga gcaactgcgc ctgaaacacc tgggtccactg
24601 tcgccgccac aagtgctttg ccgcgcgact cgggtgagtt tgcttacttt gaattgcccg
24661 aggatcatat cgagggcccc gcgcacggcg tccggcttac cgccaggga gagcttgccc
```


Abstract

454 Life Sciences has developed proprietary methods for massively parallel DNA sequencing. We have applied this technology to re-sequencing and mapping human BAC clones to their precise chromosomal locations. This preliminary data shows the efficacy of the technology to rapidly sample and characterize subsets of sequence spanning an entire genome or a specific chromosomal location. The novel DNA sequencing method consists of three steps: template preparation, solid phase amplification, and solid phase DNA sequencing. Several thousand to several hundreds of thousands of DNA sequencing reactions are performed simultaneously on glass plates containing 300 thousand to 1 million, 75 picoliter volume wells. Average read length of each fragment is consistently greater than 50 bases. The starting point for genome sequencing involves a single template preparation and an absence of a bacterial plasmid cloning step, thus greatly reducing costs and increasing the throughput of our system. In addition, we are completing development of a new software algorithm for *de novo* whole genome assembly. Sequencing results from human BAC clones will be presented and discussed.

Experimental Procedures

Our novel methodology requires only a single sample preparation per genome, utilizes simultaneous clonal amplification of shotgun fragments in sub-nanoliter microreactors, without the use of time-consuming cloning steps. The product of each microreactor is driven to and captured by a concomitant solid support. The captured DNAs are delivered to wells on the PicoTiterPlate™ and sequenced on 454 Life Sciences' sequencing platform. The details of these steps are illustrated in Figure 1.

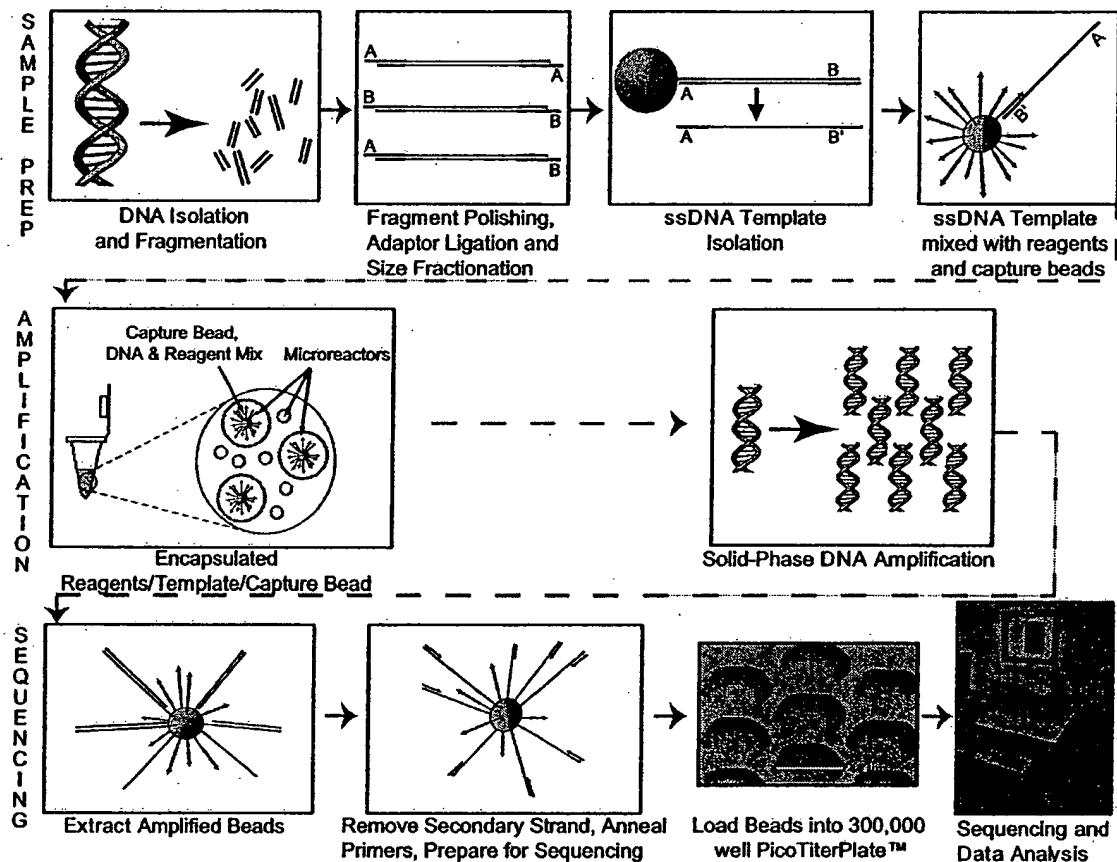


Figure 1. Streamlined template preparation and amplification process

- 1) BAC DNA from clone RP11-418C2 was fragmented to sub-kilobase lengths.
- 2) The fragment ends were polished, 5' and 3' adaptors ligated onto each fragment, and the sample was size fractionated, resulting in products under 500 bases in length.
- 3) One strand of these double-stranded products was bound to microparticles, and the free strand was eluted as template for the subsequent amplification reaction.
- 4) Amplification was conducted in a single reaction preparation, encapsulating the reaction reagent mix, a single DNA capture bead, and template in a 40 to 100 picoliter microreactor.
- 5) The particular template molecule contained in each individual microreactor was amplified and immobilized on the respective DNA capture bead.
- 6) The DNA capture beads were extracted and the template DNA was prepared for use on the 454 sequencer.

Data Analysis

The 454 sequencer generates raw traces for each microreactor, and produces sequence reads in FASTA format using a proprietary basecaller program. Adaptors and low quality reads are removed and repeats masked before mapping and assembly.

Human Genome Mapping:

Each masked read was mapped against the human genome (NCBI build 33) using BLAT and the mapped reads (>95% identity) are recorded for each chromosome.

BAC Assembly:

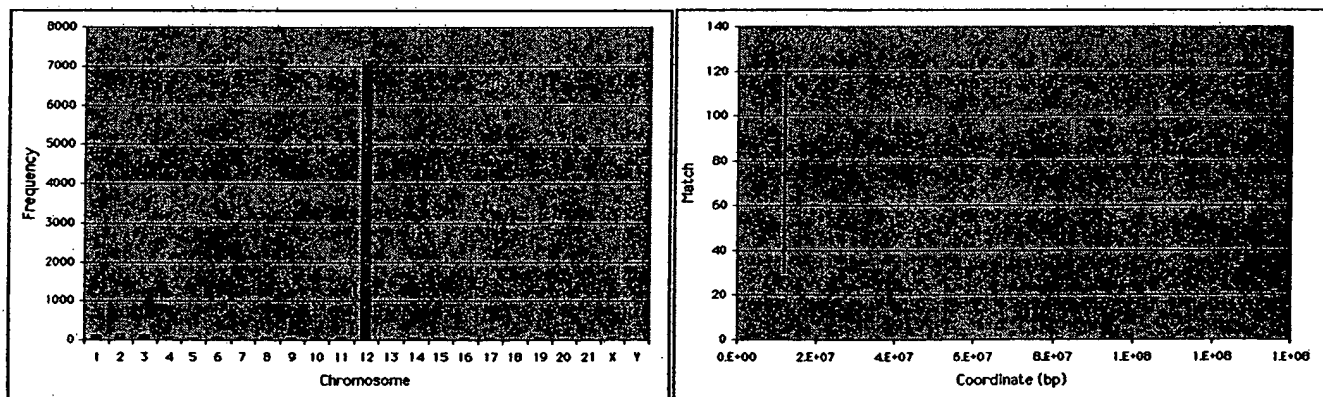
Each sequence was mapped against the reference BAC sequence (RP11-418C2) using a proprietary alignment algorithm and the resulting alignment was recorded. For sequences that map to the genome with >90% accuracy, the software generates a list of individual bases found at a given position in the reference genome. The consensus base for each location was computed by averaging all mapped bases. This consensus sequence was then compared with the reference sequence to calculate total accuracy and coverage.

We also mixed 3x oversample of reads (950 sequences) generated from conventional Sanger method with reads generated from the 454 sequencer and assembled with Phrap using default parameters.

Results and Discussion

Human Genome Mapping:

Out of 8561 mapped reads, 7153 are mapping to human chromosome 12 (Fig. 2a). Of these, 7058 reads map to the expected location within chromosome 12 (Fig. 2b). The coordinate boundaries for clone RP11-418C2 in NCBI build 33 are 11,818,492-11,986,440, whereas boundaries on the 7058 read stack are 11,816,616-11,986,511. We also mapped these reads to the mouse genome, and located the BAC to the syntenic region on mouse chromosome 6 (data not shown).



(a) Mapping against Human Genome

(b) Mapping to Human Chr12

Figure 2. Sequence mapping against human genome and within chromosome 12

Results and Discussion

BAC Assembly:

In a separate sequencing run, we generated 67193 raw reads from this BAC clone. After adaptor removal, repeat masking and quality trimming, 39900 reads were assembled against the reference sequence (Fig. 3). Genome coverage is 85% and consensus accuracy is 98%. Average read length is 84 bases.

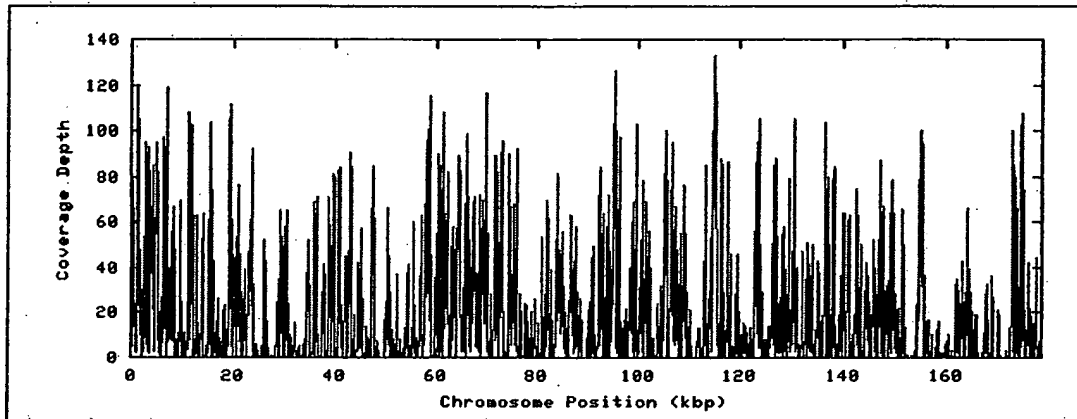


Figure 3. Frequency of assembled reads across BAC sequence length

Phrap Assembly:

Sanger reads alone generated 25 major contigs (>2 kb) with a 76% mapping efficiency, whereas Sanger and 454 reads combined produced 18 major contigs with a 83% mapping efficiency. 454 reads were able to join and extend Sanger contigs into much larger stretches (Fig. 4).

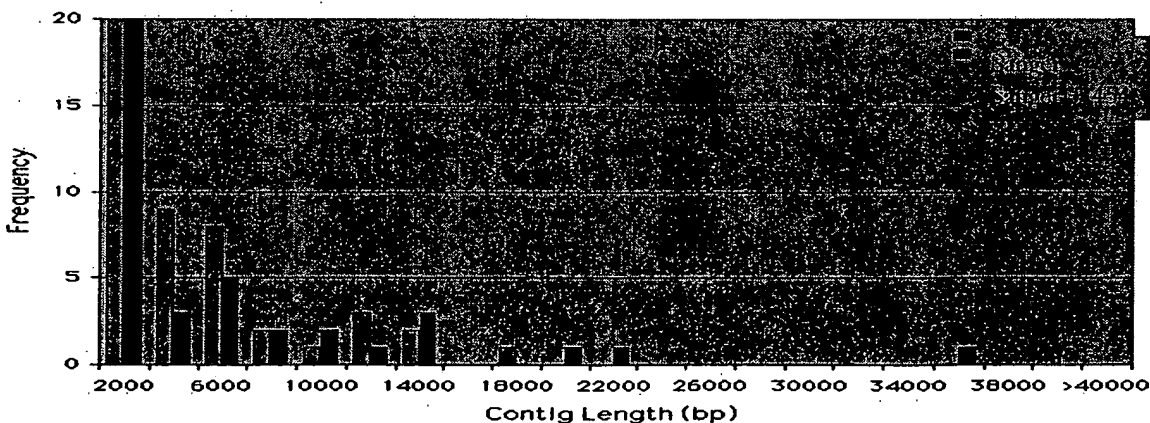


Figure 4. Distribution of Phrap contigs from Sanger only and Sanger+454 reads

Conclusion

We have demonstrated in this study that 454 Life Sciences' novel sequencing methodology is capable of producing sufficient shotgun sequence coverage of a BAC clone in a single run (done within 1 day). The reads can be used to map its precise location in the genome, as well as assembling into contigs based on a reference sequence. This is a useful tool for whole genome mapping and sequencing.

We also showed that by combining conventional Sanger method with 454 technology, we achieve a better *de novo* assembly outcome for whole genome shotgun sequencing.

We are continuing to develop our quality scoring and trimming algorithm. We have completed phase one of our proprietary fragment assembler, designed to take advantage of the raw trace signals produced by our sequencing-by-synthesis method. This assembler will be available as part of 454's commercial sequencing instrument.